X. PROCESSING AND TRANSMISSION OF INFORMATION

Prof. E. Arthurs
Prof. P. Elias
Prof. R. M. Fano
Prof. J. Granlund
Prof. D. A. Huffman
Prof. H. Rogers, Jr.
Prof. C. E. Shannon (absent)
Prof. W. M. Siebert
Prof. J. M. Wozencraft
R. K. Brayton
D. C. Coll
J. E. Cunningham
M. A. Epstein
E. F. Ferretti
R. G. Gallager
T. L. Grettenberg
P. R. Hall
F. C. Hennie III
E. M. Hofstetter
T. Kailath
D. C. Luckham
R. S. Marcus
J. Max
G. E. Mongold, Jr.
G. E. W. Sparrendahl (visitor)
W. A. Youngblood
H. L. Yudkin

A. STUDIES OF WOODWARD'S UNCERTAINTY FUNCTION

The function

$$\theta(\tau,\omega) = \int_{-\infty}^{\infty} u(t-\tau/2)\, u^*(t+\tau/2)\, \exp(-j\omega t)\, dt \tag{1}$$

was introduced by Ville (1), and its significance with respect to the quality of radar measurements was suggested by Woodward (2), and enlarged upon by Siebert (3). For some time we have been studying the mathematical properties of this function – with the hope of ultimately devising the necessary and sufficient conditions that a function of two variables be representable as in Eq. 1. Such conditions might prove useful in evolving a theory of radar synthesis. The results obtained thus far are summarized below. Proofs are presented only when the method of proof is not obvious.

Definitions

1. We assume that u(t) is a reasonably well-behaved, complex-valued function of the real variable, t. In particular, any integrals involving u(t) are assumed to exist.

2. We define

$$U^*(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(t)\, \exp(-j\omega t)\, dt \tag{2}$$

so that

$$u(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U^*(\omega)\, \exp(j\omega t)\, d\omega \tag{3}$$

3. We shall call a complex function $\theta(\tau,\omega)$ of two real variables, τ and ω, a θ-function if and only if there exists a function u(t) which is such that $\theta(\tau,\omega)$ may be represented as in Eq. 1.

4. We shall call a real positive function $\psi(\tau,\omega)$ of two real variables, τ and ω, a

ψ-function if and only if there exists a function u(t) which is such that

$$\psi(\tau,\omega) = |\theta(\tau,\omega)|^2 = \left|\int u\left(t-\frac{\tau}{2}\right)u^*\left(t+\frac{\tau}{2}\right)\exp(-j\omega t)\,dt\right|^2 \tag{4}$$

Theorems

1. $\theta(\tau,\omega) = \theta^*(-\tau,-\omega)$
2. If $\theta(\tau,\omega)$ is a θ-function, it has the additional representations

$$\theta(\tau,\omega) = \int_{-\infty}^{\infty} U\left(\mu-\frac{\omega}{2}\right)U^*\left(\mu+\frac{\omega}{2}\right)\exp(-j\mu\tau)\,d\mu \tag{5}$$

and

$$\theta(\tau,\omega) = \frac{1}{(2\pi)^{1/2}}\iint_{-\infty}^{\infty} u(\rho-\tau)\,U(\mu-\omega)\,\exp(-j\mu\rho)\,\exp[(j\omega\tau)/2]\,d\mu\,d\rho \tag{6}$$

3. If $\theta(\tau,\omega)$ is a θ-function corresponding to u(t), then $(1/a)\left(\theta[a\tau,(\omega/a)]\right)$ is a θ-function corresponding to $u(at)$.

4. If $\theta(\tau,\omega)$ is a θ-function corresponding to u(t), then $\theta(\tau,\omega+2k\tau)$ is a θ-function corresponding to $\exp[jkt^2]\,u(t)$.

5. If $\theta(\tau,\omega)$ is a θ-function corresponding to $U(\omega)$, then $\theta(\tau+2a\omega,\omega)$ is a θ-function corresponding to $\exp[ja\omega^2]\,U(\omega)$.

6. If $\theta(\tau,\omega)$ is a θ-function corresponding to u(t), then $\cos\phi\,\theta(\omega\sin\phi+\tau\cos\phi,\ \omega\cos\phi-\tau\sin\phi)$ is a θ-function corresponding to the time function

$$\exp\left(-j\frac{t^2\tan\phi}{2}\right)\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} U^*(\omega)\exp\left(-j\frac{\omega^2\tan\phi}{2}\right)\exp\left(j\frac{\omega t}{\cos\phi}\right)d\omega$$

In other words, the property of being a θ-function is independent of a rotation in the coordinate axes.

7. If $\theta(\tau,\omega)$ is a θ-function, then along any straight line through the origin $\{[\theta(\tau,\omega)]/[\theta(0,0)]\}$ has the properties of a characteristic function; for example,

$$\int_{-\infty}^{\infty}\theta(\tau\cos\phi,\ -\tau\sin\phi)\,e^{-j\omega\tau}\,d\tau \geq 0 \tag{7}$$

for all τ and ω.

8. A necessary and sufficient condition that $\theta(\tau,\omega)$ be a θ-function

is that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \theta(\tau,\omega-\mu)\exp\left(j\frac{(\omega+\mu)\tau}{2}\right)d\tau \tag{8}$$

shall factor in the form $f(\mu)\,f^*(\omega)$. If this condition is satisfied, then $U(\omega)$ can be identified with $f(\omega)$.

Proof:

Necessity follows directly upon substituting Eq. 5 into Eq. 8. To prove sufficiency, assume that Eq. 8 factors and set $\mu = \phi - (\rho/2)$ and $\omega = \phi + (\rho/2)$. Then we have

$$\int_{-\infty}^{\infty} f\left(\phi-\frac{\rho}{2}\right) f^*\left(\phi+\frac{\rho}{2}\right)\exp(-j\phi\xi)\,d\phi = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta(\tau,\rho)\exp(j\phi\tau)\,d\tau\right]\exp(-j\phi\xi)\,d\phi$$

$$= \theta(\xi,\rho)$$

which is valid if $f(\omega)$ is identified with $U(\omega)$.

9. An equivalent necessary and sufficient condition that $\theta(\tau,\omega)$ be a θ-function is that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \theta(\tau-p,\omega)\exp\left(j\frac{(\tau+p)\omega}{2}\right)d\omega \tag{9}$$

factor in the form

$$f(p)f^*(\mu)$$

If this condition is satisfied then $u(t)$ may be identified with $f(t)$.

10. If $\theta_1(\tau,\omega)$ and $\theta_2(\tau,\omega)$ are both θ-functions and neither is identically zero, then $\theta(\tau,\omega) = \theta_1(\tau,\omega) + \theta_2(\tau,\omega)$ is a θ-function if and only if $\theta_1(\tau,\omega) = C\theta_2(\tau,\omega)$, where C is a constant.

Proof:

The sufficiency of the condition is obvious. Necessity follows from Eq. 8 because we must have $U(\mu)\,U^*(\omega) = U_1(\mu)\,U_1^*(\omega) + U_2(\mu)\,U_2^*(\omega)$ for all μ and ω. It is easily shown that this can be true only if $U_1(\mu)$ is proportional to $U_2(\mu)$; that is, if $\theta_1(\tau,\omega)$ is proportional to $\theta_2(\tau,\omega)$.

11. If $\theta_1(\tau,\omega)$ and $\theta_2(\tau,\omega)$ are both θ-functions, then both

$$\theta'(\tau,\omega) = \int_{-\infty}^{\infty}\theta_1(t,\omega)\,\theta_2(\tau-t,\omega)\,dt \tag{10}$$

and

$$\theta''(\tau, \omega) = \int_{-\infty}^{\infty} \theta_1(\tau, \mu)\, \theta_2(\tau, \omega - \mu)\, d\mu \tag{11}$$

are also θ-functions. In the case of (10)

$$U'(\omega) = U_1(\omega)\, U_2(\omega)$$

and in the case of (11)

$$u''(t) = u_1(t)\, u_2(t)$$

Theorems 3, 4, 5, and 6 also apply, with obvious modifications, to ψ-functions in place of θ-functions. In particular, the property of being a ψ-function is independent of a rotation in axes. Other theorems are:

12. If $\psi(\tau, \omega)$ is a ψ-function, then $\psi(\tau, \omega)$ is its own two-dimensional Fourier transform, i.e.,

$$\frac{1}{2\pi} \iint_{-\infty}^{\infty} \psi(\tau, \omega)\, e^{-j\mu\tau}\, e^{j\omega p}\, d\tau\, d\omega = \psi(p, \mu) \tag{12}$$

13. If $\psi(\tau, \omega)$ is a ψ-function, then

$$\frac{1}{2\pi} \iint_{-\infty}^{\infty} \psi(\tau, \omega)\, d\tau\, d\omega = \psi(0, 0) \geq \psi(\tau, \omega) \tag{13}$$

14. If $\psi_1(\tau, \omega)$ and $\psi_2(\tau, \omega)$ are both ψ-functions, then

$$\psi(\tau, \omega) = \psi_1(\tau, \omega) + \psi_2(\tau, \omega)$$

is a ψ-function if and only if $\psi_1(\tau, \omega) = C\psi_2(\tau, \omega)$ where C is a real constant, ≥ -1. The proof depends on arguments similar to those involved in Theorem 10. The significance of this theorem is that it proves that the condition of Theorem 12 is not sufficient.

15. If $\psi(\tau, \omega)$ satisfies condition 12 and _a fortiori_ if $\psi(\tau, \omega)$ is a ψ-function

$$\iiiint_{-\infty}^{\infty} \psi(\tau_2 - \tau_1, \omega_2 - \omega_1)\, g(\tau_1, \omega_1)\, g^*(\tau_2, \omega_2)\, d\tau_1\, d\tau_2\, d\omega_1\, d\omega_2 \geq 0$$

for any function $g(\tau, \omega)$.

16. If $\psi_1(\tau, \omega)$ and $\psi_2(\tau, \omega)$ satisfy condition 12 and _a fortiori_ if $\psi_1(\tau, \omega)$ or $\psi_2(\tau, \omega)$ are ψ-functions, then

$$\psi(\tau,\omega) = \int_{-\infty}^{\infty} \psi_1(t,\omega)\ \psi_2(t-\tau,\omega)\ dt$$

also satisfies condition 12 but is not necessarily a ψ-function.

Necessary and sufficient conditions for $\psi(\tau,\omega)$ to be a ψ-function have not yet been discovered.

W. M. Siebert

References

1. J. Ville, Théorie et application de la notion de signal analytique, Câbles et Transmission, Vol. 2, no. 1, pp. 61-74 (Jan. 1948).
2. P. M. Woodward and I. L. Davies, A theory of radar information, Phil. Mag. 41, 1001-1017 (1950).
3. W. M. Siebert, A radar detection philosophy, Trans. IRE, vol. IT-2, no. 3, pp. 204-221 (Sept. 1956).

B. CODING THEOREM FOR FINITE-STATE CHANNELS

Proof will be given that finite-state channels obey a slightly weakened version of the fundamental theorem for noisy channels (1). First, we define a finite-state channel (see Fig. X-1) by the following properties:

1. At any time the channel is in one of a finite number of states;

2. The state transitions and outputs are governed by a fixed set of probabilities, $\Pr\left(y_i^o s_j^o \mid x_k^o s_\ell^{-1}\right)$, defined for each output, $y_i \in Y$; for each input, $x_k \in X$; and for each pair of states, s_j, $s_\ell \in S$. (Throughout the paper, superscripts denote time; subscripts denote particular letters.)

The simplest example of a finite-state channel is a channel in which both the transmitter and receiver know the channel state. We can show that the maximum information through this type of channel can be achieved by sources in which the probabilities of the

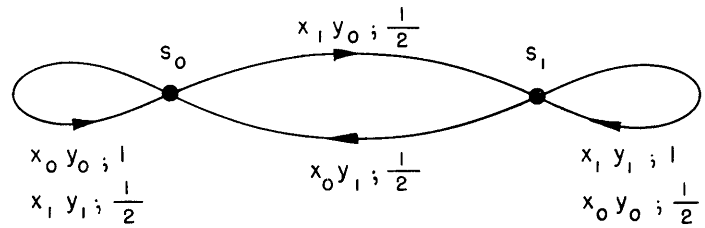

Fig. X-1. Finite-state channel ("billiard-ball"). The third number on each transition line is $\Pr\left(y_i^o s_j^o \mid x_k^o s_\ell^{-1}\right)$.

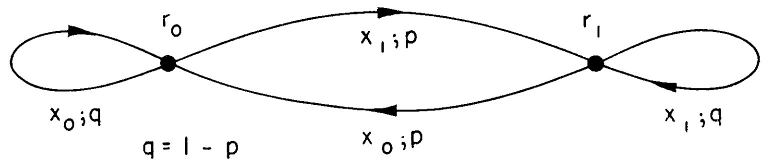

Fig. X-2. Finite-state source. The second letter on each transition is $\Pr\left(x_i^o r_j^o \mid r_k^{-1}\right)$.

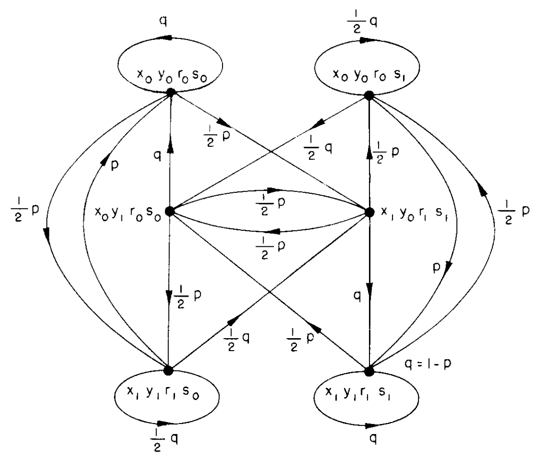

Fig. X-3. Markov diagram combining Figs. X-1 and X-2.

input letters are dependent only upon the channel state. Since these channels can be handled in much the same way as memoryless channels, they will not be considered here.

In the next case that we consider, the transmitter knows the channel state, but the receiver does not. It would seem reasonable that channel capacity could be reached by a source with probabilities that depend only upon the state. But this is not so. Heuristically, the reason is that in order to reach channel capacity the transmitter should partially base his strategy on the condition of the receiver. However, the receiver's estimate of the state is based on all of the past outputs. Since this case is essentially no simpler than the case in which neither the transmitter nor the receiver knows the state, we shall consider the general case.

In the general finite-state channel, there can exist statistical constraints that extend through all time. An example is found in a channel that changes state on each even-numbered occurrence of the input x_1. In cases like this, we would not expect to approach channel capacity merely by introducing into the source statistical constraints that extend farther and farther back into time. However, it seems entirely reasonable (although unproven) that we can approach channel capacity by using finite-state sources with more and more states. A finite-state source, like a finite-state channel, is given by the set of probabilities, $\Pr\left(x_i^o r_j^o \mid r_k^{-1}\right)$, defined for each input, $x_i \in X$; and for each pair of source states, $r_j, r_k \in R$. (See Fig. X-2.)

From this reasoning, we are led to define the channel capacity, C, of a finite-state channel as the lowest upper bound of the average mutual information rate, taken over all possible finite-state sources that can be connected to the channel. This definition allows us to prove the following theorem.

THEOREM. Given any $\epsilon > 0$ and $\delta > 0$, and given any finite-state channel with capacity C, there exist codes that transmit information over the channel at rates greater than $C - \epsilon$ with a probability of decoding error, Pr(e), less than δ.

PROOF. We can find a finite-state source that transmits through the channel at a rate, $\bar{I} \geq C - (\epsilon/4)$. With this source connected to the channel, we can consider each quadruplet of input, output, source state, and channel state – that is, (x_i, y_j, r_k, s_ℓ) – as a superstate of a Markov chain and designate it by $w_{ijk\ell}$. (See Fig. X-3.) Clearly the source and channel transition probabilities define the transition probabilities on the Markov chain. This Markov chain has a finite number of states and in general (2) can be broken up into two parts:

1. One or more closed sets of states. (A closed set is a set from which no transitions are possible to states outside of the set.)
2. Possibly some transient states.

We assume that the source and channel are operating in one of the closed sets of states in which the associated average mutual information rate through the channel is $\bar{I} \geq C - (\epsilon/4)$. We can now forget all of the other states and consider only this one irreducible chain. There are two possible cases: (a) The chain is ergodic; (b) the chain is periodic. The proof will be carried through for case (a). Then we shall show that the proof is essentially the same for case (b).

First, we shall prove that the output from the channel is ergodic. The occurrence of any particular Y sequence is a function of the state sequences on the super-Markov graph. Therefore, by the ergodic theorem, the ensemble probability of occurrence of any Y sequence is equal to its relative frequency of occurrence in any infinite length sequence except in a set of sequences of zero probability. It can be shown (3) that this condition is sufficient to guarantee the ergodicity of the output sequence. In exactly the same way it can be established that the input is ergodic and that the combination of input and output is ergodic.

Because of the ergodicity of the output, it can be shown (4) that

$$\lim_{n\to\infty} \frac{1}{n}\left[-\log_2 \Pr(y^1 y^2 \ldots y^n) - H(Y^1 Y^2 \ldots Y^n)\right] = 0 \tag{1}$$

for all sequences except a set of zero probability. Likewise,

$$\lim_{n\to\infty} \frac{1}{n}\left[-\log_2 \Pr(x^1x^2\dots x^n) - H(X^1X^2\dots X^n)\right] = 0 \tag{2}$$

$$\lim_{n\to\infty} \frac{1}{n}[-\log_2 \Pr(x^1y^1\dots x^ny^n) - H(X^1Y^1\dots X^nY^n)] = 0 \tag{3}$$

except on a set of zero probability.

Combining these three statements, we now show that for sufficiently long sequences the mutual information in most sequences tends to the average mutual information. The average mutual information, $\overline{I}$, is given by

$$\lim_{n\to\infty} \frac{1}{n}[H(X^1\dots X^n) + H(Y^1\dots Y^n) - H(X^1Y^1\dots X^nY^n)] \tag{4}$$

The mutual information rate, I_n, in a particular sequence of length n is given by

$$\frac{1}{n}\left[\log_2 \Pr(x^1y^1\dots x^ny^n) - \log_2 \Pr(x^1\dots x^n) - \log_2 \Pr(y^1\dots y^n)\right] \tag{5}$$

Using Eqs. 1, 2, and 3, we see that $\lim I_n = \overline{I}$ except for a set of sequences of zero probability.

We conclude the proof by using a theorem proved by Shannon (5). For the particular source that we have chosen, I_n is a random variable which is defined by Eq. 5 for every sequence of X and Y of length n. Then a distribution function can be defined:

$$\rho(x) = \Pr(I_n < x)$$

Shannon's theorem then states that for any integer M and any $\theta > 0$, there exists a selection of M input words of length n that is such that if the M words are used with equal probability as the input to the channel, then the probability of a decoding error, Pr(e), is bounded by

$$\Pr(e) \leqslant \rho\left(\frac{1}{n}\log_2 M + \theta\right) + 2^{-n\theta} \tag{6}$$

This rather remarkable theorem is proved by considering an ensemble of codes in which each code word is chosen with the probability assigned by the particular finite-state source that was chosen at the beginning of the proof.

Although Shannon's theorem is proved in a paper concerned with memoryless channels, the restriction to memoryless channels is used only in assuming that successive blocks are independent of each other. In finite-state channels, the theorem still holds if up until time zero the finite-state source is used and then one block of code is transmitted. If more than one block is transmitted, the use of the code on one block will change the state probabilities for the beginning of the next block and hence will change all of the I_n. However, we are interested only in the effect on Pr(e). The worst situation that can occur is a combination of the following events:

1. All of the decoding errors occur when the channel starts in state s_i.

2. $\Pr(s_i) = [\Pr(s_j)]_{min}$, in which these are the probabilities of the states with the finite-state source attached to the channel.

3. The code that satisfies Eq. 6 for one block will always leave the channel in state s_i. Even for these extreme conditions, we have

$$\Pr(e|s_i)\Pr(s_i) = \Pr(e)$$

$$\Pr(e|s_i) = \Pr'(e) = \text{probability of error when the code is used on successive blocks}$$

Hence, in general, we have the result

$$\Pr'(e) \leqslant \frac{\rho\left(\frac{1}{n}\log_2 M + \theta\right) + 2^{-n\theta}}{[\Pr(s_j)]_{min}}$$

Since $[\Pr(s_j)]_{min}$ is a function only of the channel and the selected finite-state source, it is independent of n. For any n sufficiently large, we can find an M which is such that

$$C - \epsilon < \frac{\log_2 M}{n} \leqslant C - \frac{3}{4}\epsilon \tag{7}$$

in which C is the channel capacity, previously defined as the lowest upper bound of the channel information rate using finite-state sources. Choose $\theta = (\epsilon/4)$. Then

$$\rho\left(\frac{1}{n}\log_2 M + \theta\right) \leqslant \rho\left(C - \frac{\epsilon}{2}\right)$$

Since $\lim\limits_{n\to\infty} I_n = \overline{I} \geqslant C - (\epsilon/4)$ with probability 1, we have

$$\lim_{n\to\infty} \rho\left(C - \frac{\epsilon}{2}\right) = 0$$

$$\lim_{n\to\infty} 2^{-n\theta} = 0$$

$$\lim_{n\to\infty} \Pr'(e) = 0$$

so that for n sufficiently large, the probability of decoding error is less than δ and, from Eq. 7, the rate is greater than $C - \epsilon$. This concludes the proof of case (a).

To prove case (b), let us assume that the Markov chain has a period k. We can form an ergodic Markov chain from the periodic one by combining sequences of k states into one state on the new graph. By considering only input and output sequences whose lengths are multiples of k, we can prove the theorem exactly as before. Q.E.D.

The results of this theorem were applied to the "billiard-ball channel," a particularly simple example of finite-state channels that is described in Fig. X-1. It was formerly thought that the capacity of this channel was 0.5 bits, achieved by sending each input letter twice in succession. However, we found a finite-state source that signalled through the channel at a rate of 0.554, and by the theorem, this is achievable with arbitrarily small error.

It is conceivable that nonfinite-state sources will signal at a rate greater than C over certain finite-state channels. But note that no block code can possibly signal at a rate greater than C, since block codes are a class of finite-state sources.

Finally, attention should be called to the technique of forming a super-Markov chain from a finite-state source and finite-state channel. This promises to be a useful conceptual tool in further work on finite-state sources and channels.

R. G. Gallager

References

1. C. E. Shannon, The Mathematical Theory of Information (University of Illinois Press, Urbana, 1949).
2. W. Feller, Probability Theory and its Applications (John Wiley and Sons, New York, 1950) Chap. 15.
3. A. I. Khinchin, Mathematical Foundations of Information Theory (Dover Publications, Inc., New York, 1957) Chap. 4.
4. B. McMillan, The Basic Theorems of Information Theory, Bell System Monograph 2124 (1953).
5. C. E. Shannon, Certain results in coding theory for noisy channels, Information and Control (Sept. 1957).